

УДК 004.8

ИНФОРМАЦИОННЫЙ ПОДХОД В ЭТИКЕ*

*Фоминых И.Б., д.т.н., профессор,
ФГБОУ ВПО «НИУ «МЭИ»
e-mail: igborfomin@mail.ru*

1. ВВЕДЕНИЕ

В докладе рассмотрены некоторые проблемы развития информационного подхода [1,2] в этике [3,4], тесно связанные с проблемами эмоций и эмоциональной оценки агентами друг друга. В ранее опубликованных работах [5-7] было введено понятие алгебры эмоций, основанной на оценках. Алгебра эмоций позволяет вычислять величину и знак сложной эмоции с помощью простых операций над ее составляющими. Эта алгебра существует объективно как существенная сторона самого механизма эмоций, изначально предназначенного именно для количественных оценок. Результатом вычислений является величина эмоции и её знак - положительный или отрицательный.

Алгебра эмоций, судя по некоторым фактам, реализуется в мозгу человека в аналоговой форме с помощью нейроподобных распределенных вычислительных структур, задача которых оперировать оценками (причем, как в области материальных благ, так и благ нематериальных: поступков, мыслей, чувств и т.д.).

В общем случае оценка c_{ij} вводится как отношение:

$$C_{ij} = dx_i/dx_j,$$

где dx_i , dx_j - приращение значений переменных x_i , x_j .

Тогда оценка - это относительная ценность единицы dx_j , выраженная в единицах dx_i . Например, если dx_j - услуги, а dx_i - деньги, то C_{ij} - цена единицы услуги, выраженная в деньгах.

С этой точки зрения эмоция e ,

$$e = dL_i/dt = dL_i/dx_i dx_j/dt,$$

есть оценка единицы времени dt в единицах такой переменной как целевая функция dL .

*Работа выполнена при финансовой поддержке РФФИ (проекты №13-07- 00396, 14-07-00373, 15-07-02320)

2. ОЦЕНКИ СЛОЖНЫХ ЭМОЦИЙ

Как показано в работе [7], на базе двух простейших эмоций (удовольствие и страдание) для двух субъектов, можно ввести еще четыре, названные условно сорадость, зависть, сочувствие и злорадство. Если в систему оценок ввести третьего субъекта k , который наблюдает за субъектами j и i и дает оценку чувствам последнего, то формула такой оценки e_{kj} может быть представлена в виде:

$$e_{kj} = c_{kj} c_{ji} e_i,$$

где c_{kj} - оценка субъекта j субъектом k ; c_{ji} - оценка субъекта i субъектом).

Пусть этим субъектом будет Абсолютный Авторитет (Абсолют), оценки которого признаются обязательными для всех членов группы. Представление об Абсолюте в психике субъекта играет роль его нравственного сознания, совести, «сверх-Я». Оно задаст некоторое абсолютное начало отсчета для оценок. Оценка, даваемая субъектом k субъекту j , является относительной и зависит от того, хорош или плох сам субъект k . Конечно, априори каждый склонен считать сам себя хорошим, пока ему не доказали обратное («презумпция позитивности»). Однако другие субъекты могут иметь на этот счет свое мнение. Наличие единого, всеми признаваемого Абсолюта приводит к ликвидации различий в оценках: отношение \wedge авторитетной инстанции к субъекту i , может рассматриваться как абсолютная оценка субъекта i как «плохого» или «хорошего». Но и отношение c_{kj} субъекта к Абсолюта к j будет характеризовать теперь уже не столько этот Абсолют (он позитивен по определению), сколько самого субъекта k . Кто хорошо относится к авторитету, тот субъект хороший, а кто плохо - тот сам плохой.

Рассмотрим теперь следующий случай: субъект i испытывает страдание e_i^- , а субъект k - радуется по этому поводу, субъект j - Абсолют. Это можно записать в виде следующего соотношения:

$$e_{kj}^+ = c_{kj} c_{ji} e_i^- \quad (1)$$

Напомним, что знак оценки пишется справа и сверху от нее. Знак произведения определяется по известным правилам алгебры.

Знаки оценок c_{kj} и c_{ji} в (1) намеренно оставлены неопределенными. Однако из уравнения следует, что произведение этих оценок должно быть отрицательным. Возможны два случая: либо субъект k - положительный человек (c_{kj}^+), а субъект i - плохой (c_{ji}^-), заслуживший свое несчастье, так что злорадство субъекта k оправдано.

Либо же субъект i - хороший человек (c_{ji}^+), но сам субъект k - плохой (c_{kj}^-), и тогда понятно, почему он радуется несчастью хорошего человека. Таким образом, если заданы три члена уравнения, то естественно всегда можно определить четвертый. Этим неизвестным может быть как эмоция, так и оценка одного из субъектов. Таким образом, формула для сложных эмоций позволяет не только вычислять эмоции, если заданы оценки субъектов, но и «вычислять» оценки самих субъектов как положительных или отрицательных, если заданы их эмоции.

Особый класс эмоций возникает, когда субъект и объект - одно и то же лицо. На самом деле субъект и объект, конечно, совпадать не могут, речь идет об оценке одного уровня личности другим уровнем (обычно низшего высшим). Согласно принятым представлениям (см., например, [4]), человек имеет три уровня организации: физическую природу («Оно»), одушевленный уровень («Я») и, наконец, совесть («Сверх-Я»). Этот верхний уровень часто обладает целевой функцией, отличной от целевой функции «Я». Таким образом, субъект k может радоваться несчастью субъекта i , но в то же время, его совесть (внутренний субъект) будет осуждать его за это.

Часто возникает и более общий вопрос об оценке животного начала человека (k) со стороны его духовного начала (j).

3. ЭТИКА И ЭМОЦИИ

Воспользуемся введенным аппаратом оценок для исследования соотношений между этикой языческой (свойственной в значительной степени и обыденному сознанию) и этикой христианской. Как известно, христианская этика отрицательно относится к животной природе человека. Корни этого отношения уходят в мир физический. Физические законы сохранения материи и энергии приводят к тому, что «сколько чего в одном месте убавится, столько же в другом присовокупится». И если материальные ресурсы дефицитны, то приобретения одного оборачиваются потерей для другого. Поэтому со времен Аристотеля, материальное всегда было этически подозрительным как потенциальный источник соперничества и вражды. Иное дело духовное начало, информация. Для нее закон сохранения отсутствует. Информацией можно делиться, ничего не теряя: если в одном месте присовокупится, то в другом ничего не убавится. Здесь понятны корни негативного отношения к материальному, корни аскетизма.

Аскетизм проповедовали многие этические системы (буддизм, манихейство, стоицизм). Но только христианство сделало следующий

логически неизбежный шаг, выводящий эту идею из плана сугубо личного в план межличностный: «будь последователен в пренебрежении к своему материальному - возлюби того, кто наносит ему ущерб: врага своего». В языческой этике существует порочный круг- зла - положительная обратная связь между негативными эмоциями: «око за око, зуб за зуб». Примеров того, как раскручивается эта зловещая спираль-великое множество. Во введенной ранее системе понятий этот процесс описывается следующим уравнением:

$$c_{ki} \, dc_{ik}/dt = c_{ki} (\partial L^i/\partial L^k - c_{ik}) > 0,$$

где ∂L^i , ∂L^k - приращения целевых функций за время dt субъектов i , k соответственно.

Если c_{ki} отрицательно, то и скорость изменения dc_{ik}/dt , согласно принципу максимума информации [1], также должна быть отрицательна. Это означает, что раз k видит в i своего врага, то он стремится действовать так, чтобы $\partial L^i/\partial L^k$ было отрицательным, т.е. будет своим действием стараться принести ущерб субъекту i . Причем эмоция его будет положительной только в том случае, если эффект его действия $\partial L^i/\partial L^k$ превзойдет (но абсолютной величине) c_{ik} - средний уровень его прежних “достижений”.

Так же действует субъект i по отношению к субъекту k . Процесс идет по нарастающей. Даже если начальные значения c^* и c_{ki} были ничтожно малыми, но негативными, они быстро возрастают под действием положительной обратной связи, пока не достигают предельно возможных значений.

Таким образом, вражда обладает устойчивостью и способностью к самоподдержанию. Начавшись часто с пустяка, она быстро достигает крайних форм и удерживается в этом состоянии благодаря механизму положительной обратной связи, приводящей систему в крайнее состояние. Хорошо известные примеры - ссора Ивана Ивановича с Иваном Никифоровичем или кровная месть, которая становится несчастьем для двух враждебных кланов на протяжении многих поколений, когда первоначальная причина вражды (часто ничтожная) давно уже забылась.

Христианская этика позволяет разорвать этот порочный круг. Для этого надо одному из субъектов (например, k) преодолеть порог, принудительно изменить отрицательное значение c_{ki} на положительное. Сделать это в одиночку достаточно сложно, но намного легче, если обратиться к посредничеству Абсолюта, для которого вес люди одинаково добры. Для этого нужно сначала самому принять этот авторитет (c_{kj}^+) и принять его оценку другого субъекта (c_{kj}^+). Тогда оценка: $c_{kj} = c_{kj}^+ c_{ji}^+$ становится положительной.

С этого момента в циклическом процессе вражды возникает противоречие: оценка c_{ki} положительна, встречная оценка c_{ik} пока еще остается отрицательной. Обратная связь в целом становится уже не положительной, а отрицательной. В такой системе процесс уже не идет «в разнос», а становится устойчивым, негативные эффекты уменьшаются. Вражда постепенно затихает, если один из партнеров не отвечает злом на зло. С какого-то момента и субъект i может сменить отношение к субъекту k на позитивное (c_{kj}^+), хотя это не обязательно. Но если это случилось, то обратная связь снова становится положительной ($c_{ik} c_{ki} > 0$), однако происходит это уже при других, положительных знаках обоих отношений. И тогда начинается новый неустойчивый процесс, но уже идущий в другом направлении: цепная реакция добра.

Принятие Абсолюта требует определенного самопринуждения для преодоления порога. Но затем новое состояние становится устойчивым и способным к самоподдержанию благодаря положительной обратной связи, которая работает теперь в сторону добра. Рассмотрим некоторые следствия этого основного постулата христианской этики.

4. ПАРАДОКСЫ ХРИСТИАНСКОЙ ЭТИКИ

Христианская этика осуждает злорадство как таковое, к кому бы оно ни относилось. Это значит, что христианин (c_{kj}^+) не вправе радоваться несчастью другого, кем бы ни был этот другой. Его эмоция по поводу несчастья другого может быть только отрицательной. Расставляя соответствующие знаки в формуле (1), получаем:

$$e_{kji} = c_{kj}^+ c_{ji} e_i^-$$

Это означает, что христианин (c_{kj}^+) всегда скорбит (e_{kji}) при встрече с чужим несчастьем (e_i^-). Отсюда следует важный вывод: оставшийся неопределенным знак множителя c_{ji} , характеризующий отношение Абсолюта к субъекту i , должен быть положительным: «все люди добры» с точки зрения Абсолюта.

С позиций христианской этики оценка животного начала в человеке отрицательна (c_{ji}^-). Поэтому страдания, испытываемые нашим телом (e_i^-), должны оцениваться нашим «Я» как благо (разумеется, если само наше «Я» приемлет Абсолют и дает ему положительную оценку c_{kj}^+):

$$e_{kji}^+ = c_{kj}^+ c_{ji}^- e_i^- \quad (2)$$

Физическое страдание христианская этика претворяет в духовное наслаждение e_{kj}^+ - в силу отрицательного знака произведения сомножителей c_{kj}^+ c_{ji}^- .

И наоборот, всякое плотское наслаждение e_i^+ получает в рамках христианской этики отрицательную оценку:

$$e_{kj}^- = c_{kj}^+ c_{ji}^- e_i^+ \quad (3)$$

Иное дело в языческом мировоззрении, где физическая природа человека с точки зрения высшего авторитета оценивается положительно (сД). Заменяя это значение в формулах (2), (3), получаем, соответственно: $e_{kj}^- = c_{kj}^+ c_{ji}^+$ e_i^+ , $e_{kj}^+ = c_{kj}^+ c_{ji}^+ e_i^+$, т.е., оценка

телесных страданий меняет знак: они оцениваются, как и положено «естественному человеку» отрицательно. А наслаждения, соответственно, положительно. Отсюда и прямо противоположная оценка врага - того, кто причиняет нам телесные, физические страдания. Ненависть и месть - в языческой (обыденной) этике; любовь и прощение - в христианской.

Многим исследователям христианская этика представлялась и представляется извращением природы человека. На самом деле христианская этика лишь выявляет, идеализирует и сводит в систему нормативов потенциальные возможности, присущие именно человеку как таковому, те, что отличают человека от животного. Важнейшая из этих возможностей - человеческая способность к самопринуждению, к принудительному, вопреки эмоциям, переводу себя в другое состояние, которому сопротивляется его животная природа. Для такого перевода необходима внешняя точка опоры. И человек находит ее в слове (знаке). Только слово может стать авторитетным посредником между эмоциями (чувствами) и только в слове может быть сформулирована идея Абсолюта.

5. ЗАКЛЮЧЕНИЕ

Рассмотренный подход к оценкам сложных эмоций и вытекающих из них норм поведения может быть актуален при моделировании перспективных автономных многоагентных систем, где на первый план выдвигаются вопросы кооперации и совместных действий агентов.

Литература

1. Голицын Г.А., Петров В.М. Информация - поведение - творчество. - М.: Наука, 1991.

Фоминых И.Б. Информационный подход в этике // Сборник научных трудов VIII-й Международной научно-технической конференции «Интегрированные модели и мягкие вычисления в искусственном интеллекте» (Коломна, 18-20 мая 2015 г.). В 2-х томах. Т 1. 2015. С. 133-139.

2. Golitsyn G.A., Petrov V.M. Information and Creation. - Basel: Birkhauser Verlag, 1995.
3. Лефевр В.А. Формула человека. - М.: Прогресс, 1992.
4. Шрейдер Ю.А. Лекции по этике. - М.: МИРОС, 1994.
5. Фоминых И.Б. Адаптивные системы и информационная модель эмоций// Интеллектуальное управление: новые интеллектуальные технологии в задачах управления. Сборник научных трудов Международной конференции (ТСИТ-99, Переславль-Залесский, 6-9 декабря 1999 г.). - М.: Физматлит, 1999.
6. Фоминых И.Б. Эмоции как аппарат оценок поведения интеллектуальных систем // Сборник научных трудов X-й национальной конференции по искусственному интеллекту (КИИ-2006, Обнинск, 25-28 сентября 2006 г.) - М.: Физматлит, 2006. - Т.2. - С.687-694.
7. Фоминых И.Б., Алексеев В.В. Классификация эмоций: информационный подход // Интегрированные модели и мягкие вычисления в искусственном интеллекте. Сборник трудов IV-й Международной научно-практической конференции (Коломна, 28-30 мая 2007г.). - М.: Физматлит, 2007.